1991

Perceptual validation of an acoustic metric for measuring vowel production by deaf children

Regina M. Gilbert

Perceptual Validation of an Acoustic Metric for Measuring Vowel Production by Deaf Children

Regina M. Gilbert
Central Institute for the Deaf
St. Louis, Missouri

May 3, 1991

INTRODUCTION

Measuring and documenting the productions of profoundly deaf children has been a subject of interest as early as 1942, when Hugdins and Numbers studied production errors characteristic of profoundly hearing-impaired children. Since that time, one of the most difficult tasks has been to find a metric which accurately describes the children's productions. Several studies have used spectrographic analysis to quantify the speech of profoundly deaf children (Calvert, 1961; Angelocci, Kopp, and Holbrook, 1964; Monsen, 1974, 1976; Rothman, 1976). Other studies have used perceptual measures, such as naive and experienced listeners (Gulian, Hinds, and Nimmo-Smith, 1986), distinctive-feature analysis (Mencke and Ochsner, 1985), and intelligibility measures (Smith, 1975; Markides, 1970).

A descriptive and accurate metric is very important to the analysis of the speech sample, and to the validity of the study. This becomes especially difficult when comparing differences in speech production, such as those caused by profoundly deaf children. Many of the strategies used to measure speech of normal hearing children do not accurately account for the unusual differences in deaf children's speech. A more recent demand is that the metric must be device independent. Thus, a metric must measure the benefits afforded speech production by any of the prosthetic devices, with equal approximation. This demand was not necessary, until the recent onset of cochlear implant and tactile aid technology.

A new measurement technique of this kind would assist in tracking improvements in speech over time due to effectiveness of educational training, or maturational factors (Fourakis, Geers, Russo, and Tobey, 1991). In addition, it would provide an independent variable, in the assessment of prosthetic devices aiding profoundly hearing impaired children improve speech.

The purpose of this study is to validate the metric developed in the Geers and Moog Cochlear Implant Study which is being carried out at Central Institute for the Deaf in Saint Louis, Missouri. A perceptual study was undertaken, using speech samples which were previously collected and analyzed. A general relationship between the perceptual judgements and the assessment of productions, would confirm the effectiveness of the currently used metric. Since the ultimate objective from these devices is to improve intelligibility, it is important that the measurement technique reflects actual listeners perceptions.

Previous studies have attempted to describe which factors affect the speech of deaf children and decrease intelligibility. One of the most common findings is that fundamental and formant frequencies of vowels are restricted and neutralized in both frequency and amplitude (Angelocci et al, 1964; Rothman, 1976). One study by Smith (1975) described intelligibility of deaf children as being most greatly affected by vowel errors. As vowel errors showed the most marked and systematic decrease, intelligibility seemed to improve.

In the Angelocci, Kopp, and Holbrook (1964) study, they reported that the mean frequency and amplitude of the fundamental covered a wider range of values with profoundly deaf children compared to normal-hearing children. Conversely, normal-hearing children had a greater mean range in both frequency and amplitude for the first, second, and third formant. Angelocci et al hypothesized that for a statistically average deaf child, vowels differences were achieved by varying the frequency and amplitude of the fundamental frequency, more than the formant frequencies. Monsen (1976) amended the typically described generalization that vowel productions of the deaf tend to be neutralized and more centralized. His finding suggested that the phonological space of vowel articulation is commonly due to a restricted range of the second formant, rather than the first formant. In addition, Monsen felt that the first and second formant contributed equally to listener intelligibility.

PERCEPTUAL EVALUATION METHODS

A. *Recording Background*

The perceptual stimuli consisted of digital speech recordings of one matched group of children in the CID cochlear implant study. The three children were matched on the basis of the following criteria: chronological age, speech and language ability, aided speech perception, language, intelligence, and family support. Of these children, one was fitted with a cochlear implant (Nucleus WSP Multichannel), one with a tactile aid (Tactaid II), and one with an acoustic hearing aid (Oticon). Recordings were made of the child repeating a teachers production of one of the following words: heed, hid, head, hawed, who'd). The monosyllabic vowels (i, I, E, , u) were presented in the h_d context, and the child was asked to repeat each one five times. Two recorded sessions were used in the perceptual study, which were fifteen months apart (pre and post sessions).

The steady state of the vocalic segment was isolated and analyzed using the LPC subroutine of the commercially available ILS software. The mean first formant (F1) and the mean second formant (F2) were analyzed from 50 millisecond (ms) sections of the waveforms. (See Fourakis, Geers, Russo, and Tobey, 1991; for a complete description).

B. *Stimuli*

The taped recordings were digitized and stored on the MicroVAXII computer. The sections of the waveforms over which the mean F1 and F2 were analyzed in the production study were feathered on each side by a 20 ms Kaiser window. This produced a relatively smooth onset and offset of the energy of the stimulus. Thus, each stimulus was 90 ms long. The total number of perceptual stimuli was 298 tokens.

C. *Subjects*

There were ten subjects who served as listeners in the perceptual procedure. Five of the subjects were experienced with the speech of deaf children (teachers of the deaf at CID), while the additional five subjects were without any such experience with deaf children (inexperienced subjects). All subjects had normal hearing and were native English speakers. Since no significant differences were observed between the naive and experienced listeners, all ten were combined for this analysis.

D. *Procedure*

Each person listened to the 298 stimuli over Sennheiser headphones in a relatively quiet room. The stimuli were randomized separately for each subject. The subjects were instructed to listen to a stimulus up to a maximum of three times, and then they were asked to make a judgement about the production.

First, they were asked to identify the stimulus from a list of ten monosyllabic words each beginning with [h] and ending with [d], differing only in the vowel. The words used were *heed, hid, head, had, hod, hawed, hood, who'd, hud, heard.* Of these choices, only five vowels were actually presented as stimuli. However, the production metric used any of these ten vowels to identify the stimulus. Second, the subjects were asked to indicate a quality rating of the vowel production. On a scale of "1" to "5", the subjects defined whether they felt the stimulus was a good or poor production, in comparison to their internal standard of the vowel choice they selected. A score of "5" was described as very good, and a score of "1" was defined as very poor. The presentation of the stimuli and the collection of the responses was done by the MicroVAXII computer.

E. *Processing of the Responses*

Each subject's response to each stimulus was assigned a distance from the intended vowel by computing the Euclidean distance in logarithm F1, by logarithm F2 space from the response to the intended vowel. For example, if the intended vowel was [i], and the response was [u], the response was assigned the first and second formant values for [u], based on the Peterson and Barney data for normal-hearing children (1952). Then the distance was computed between the response and the ideal intended vowel. It was predicted that if the child's productions were improving, then the mean distance of the listener's responses from the target vowel should decrease.

RESULTS

As described earlier, the metric used in this study to assess the productions of profoundly hearing-impaired children was based on the mean Euclidean distance of the child's first and second formant frequencies from the ideal formant frequencies for the intended vowel. The perceptual measure assigned the first and second formant frequencies of the ideal to the choice made by the listeners, and the distance was computed between the formant frequencies of the listener's perceived vowel, and those of the intended vowel. By comparing the relationship between production and perception by vowel, an assessment can be made on the ability of the production metric to predict the listeners perception.

Mean Distances, Standard Deviation, and Significance

In the first three tables (Tables 1-3), the mean and standard deviation data is compared for production and perception distance over time. Table One shows the data on the cochlear implant subject. The mean distance and standard deviation for both the pre and post test conditions are presented for each vowel, in order of recording (iy= heed; ao= hawed; uw= who'd; ih=hid; eh=head). If the mean production distance decreases from the pre condition to the post condition, then the production of the child was a closer approximation of the intended vowel. Similarly, if the mean distance of the perceived vowel from the intended vowel is greater in the pre condition than in the post condition, then the listeners perceived an improvement, or a closer production of the intended vowel. T- tests were performed to test for significance of the difference in pre and post test scores. For example, the [iy] production shows a significant improvement from pre to post condition, and the listeners perceived a similar improvement. However, for [uw] the production metric indicated no significant difference between pre and post test values for either production or perception data. Tables two and three show the same data for the tactile aid and hearing aid subject, respectively.

Table four summarized the direction of change in distance from pre to post condition, and the level of significance of the difference between pre-test and post-test scores for each vowel. The direction changes of [iy] productions are matched by [iy] perceptions for all devices. Serious lack of agreement, when there is significant opposite direction changes for production and perception, occurs only once: [uw] for the hearing aid subject is significantly negative in production, but significantly positive in perception. In other instances in which production and perception change in opposite directions, one or both do not reach significance.

Correlational Studies

In order to more fully understand the relationship between production and perception, three correlational studies were conducted. The results of these studies are combined across all devices in Table five. The first column shows the correlation between the production distance and the perception distance values which were used to compute the means for Tables 1-3. The values for [eh] and [iy] are show highly significant positive correlations ($p<.0001$) between the two sets of distance scores. The correlations for [ih], [uw], and [ao] are significantly different from chance at the 0.01 level.

The second column shows the correlation between the pre-post difference scores for production and perception. The difference scores were computed by subtracting the post-test distance scores from the pre-test distance scores for production and perception. Correlations were calculated to compare the magnitude of pre-post change in the children's productions with change in the listener's perceptions. Once again, [eh] and [iy] show the strongest correlations between production and perception.

The third column is the correlation between a measure of subject agreement and the production distance measure. Subject agreement was defined as the total number of listeners who perceived the intended vowel. A negative correlation is expected between these two measures since if all ten listeners reported hearing the intended vowel, the distance value should be very small (close to the ideal vowel). Conversely, if none of the subjects correctly identified the vowel, the distance value should be much larger. All correlations are highly significant to ($p<0.0001$) with the highest correlations for [iy] and [eh].

CONCLUSIONS

This study indicated that the F1/F2 distance metric is a valid measure of speech production ability for profoundly deaf talkers. Several general arguments can be made on the basis of these results. The general agreement in change of direction is flawless for the cochlear implant subject, but has areas of opposition for the tactile aid and hearing aid subjects. An assumption may be made that this metric is best suited for evaluating cochlear implant benefits. However, to serve as a valid measure of speech production, the distance metric must be device independent. Thus, additional measures must be evaluated.

In addition, there is a certain amount of vowel partiality. The vowel sounds [iy] and [eh] seem to follow direction of change across all devices, and correlate favorably between production and perception measures. However, the sounds [ao] and [uw] do not match directional changes for the tactile aid and hearing subjects, respectively. There is also a reduced correlational relationship between the production and perception values for [ao] and [uw] when compared to [iy] and [eh]. The other vowel tested [ih], has borderline results. There is one dissimilar directional change with the hearing aid subject, and the production to perception correlations are fair.

To improve the relationship between production and perception, some alterations in the metric are recommended. A more complete formant analysis of the production measures including the amplitude and frequency of the fundamental and third formant would more accurately describe the vowels [uw] and [ao] and thus, might improve the acoustic metric for these vowels.

There are many factors that affect the accuracy of speech production measurements, and perhaps no single metric will completely describe a token. However, it is important to understand the limitations associated with the technique. Ultimately, the child must be described in terms of improvement due to a specific device, or no noticeable benefit. This perceptual study attempted to reproduce that in a laboratory setting, to measure what naive listeners find as intelligible speech of a profoundly deaf individual.

REFERENCES

Angelocci, A., Kopp, G., and Holbrook, A. (**1964**). "The vowel formant of deaf and normal hearing eleven to fourteen year old boys," *Journal of Speech and Hearing Disorders*, **29**, 156-70.

Fourakis, M., Geers, A., Russo, M., and Tobey, E. (**1991**). "Acoustic evaluation of vowels produced by hearing-impaired children," *Journal of the Acoustical Society of America Abstracts.*

Gulian, E., Hinds, P., and Nimmo-Smith, I. (**1986**). "Modifications in deaf children's vowel production: perceptual evidence," *British Journal of Audiology*, **20**, 181-94.

Hudgins, C.V. and Numbers, F.C. (**1942**). "An investigation of the intelligibility of the speech of the deaf," *Genetic Physiology*, **25**, 289-392.

Markides, A. (**1970**). "The speech of deaf and partially hearing children with special reference to factors affecting intelligibility," *British Journal of Disorders of Communication*, **5**, 126-40.

Mencke, E.O. and Ochsner, G.J. (**1985**). "Distinctive-feature analysis of the speech of deaf children," *Journal of Auditory Research*, **25**, 191-200.

Monsen, R.B. (**1974**). "Durational aspects of vowel production in the speech of deaf children," *Journal of Speech and Hearing Research*, **17**, 386-98.

Monsen, R.B. (**1976**). "Normal and reduced phonological space: The production of English vowels by deaf adolescents," *Journal of Phonetics*, **4**, 189-98.

Nober, H. (**1967**). "Articulation of the deaf," *Exceptional Children*, **33**, 611-21.

Peterson, G.E. and Barney, H.L. (**1952**). "Control methods used in a study of the vowels," *Journal of the Acoustical Society of America*, **24**, 175-84.

Rothman, H. (**1976**). "A spectrographic investigation of consonant-vowel transitions in the speech of deaf adults," *Journal of Phonetics*, **4**, 129-36.

Smith, C.R. (**1975**). "Residual hearing and speech production of deaf children," *Journal of Speech and Hearing Research*, **18**, 795-811.

Table One

Cochlear Implant Subject Data

	Production			Perception		
	Mean/ SD Pre	Mean/ SD Post	T-test	Mean/ SD Pre	Mean/ SD Post	T-test
iy	.151/.03	.095/.023	T=6.204 p=.0001	.248/.091	.089/.053	T=5.599 p=.0001
ao	.309/.079	.245/.064	T=2.95 p=.01	.278/.118	.17/.042	T=3.888 p=.001
uw	.16/.118	.178/.031	T=-.439 NS	.156/.085	.17/.089	T=-.402 NS
ih	.274/.08	.114/.066	T=6.443 p=.0001	.264/.079	.119/.086	T=5.518 p=.0001
eh	.157/.053	.165/.066	T=-.391 NS	.179/.08	.228/.078	T=-1.937 NS

Table Two

Tactile Aid Subject Data

	Production			Perception		
	Mean/ SD Pre	Mean/ SD Post	T-test	Mean/ SD Pre	Mean/ SD Post	T-test
iy	.162/.05	.165/.027	T=-0.25 NS	.276/.09	.352/.097	T=-3.357 p=.01
ao	.181/.085	.262/.131	T=-2.635 p=.01	.19/.114	.158/.086	T=.992 NS
uw	.115/.115	.125/.045	T=-.337 NS	.142/.106	.114/.074	T=1.004 NS
ih	.077/.026	.127/.041	T=-8.144 p=.0001	.107/.062	.263/.049	T=-9.111 p=.0001
eh	.121/.057	.131/.046	T=-0.45 NS	.154/.084	.152/.058	T=.081 NS

Table Three

Hearing Aid Subject Data

	Production			Perception		
	Mean/ SD Pre	Mean/ SD Post	T-test	Mean/ SD Pre	Mean/ SD Post	T-test
iy	.2/0.062	.188/0.074	T=.538 NS	.332/0.11	.229/0.079	T=2.966 p=.01
ao	.255/.049	.322/.02	T=-4.909 p=.0001	.14/.068	.189/0.051	T=-2.946 p=.01
uw	.159/.084	.267/.072	T=-7.368 p=.0001	.224/.082	.128/.076	T=4.016 p=.01
ih	.115/.033	.181/.075	T=-3.55 p=.01	.207/.096	.155/.063	T=1.836 NS
eh	.179/.046	.103/.035	T=8.098 p=.0001	.232/.079	.1/.057	T=5.909 p=.0001

Table Four

Direction of Change in Distance Metric

Pre-test to Post-test

	Production			Perception		
	CI	TA	HA	CI	TA	HA
iy	+**	-	+	+**	-*	+*
ao	+*	-*	-**	+*	+	-*
uw	-	-	-**	-	+	+*
ih	+**	-**	-*	+**	-**	+
eh	-	-	+**	-	+	+**

+ = Decrease in distance (better approximation)

- = Increase in distance (poorer approximation)

* = $p < 0.01$

** = $p < 0.0001$

Table Five

Correlations between Production and Perception for All Devices Combined

	Distance Value	Pre- Post Change in Distance	Subject Agreement
eh	0.61**	0.638*	-0.625**
iy	0.60**	0.610**	-0.667**
ih	0.423*	0.484*	-0.512**
uw	0.349*	0.242*	-0.127**
ao	0.256*	0.491*	-0.342**

* = $p < 0.01$

** = $p < 0.0001$